Real-time Face Mask Detection: Challenges and Solutions using Adaptive YOLOv3

J. Anvar Shathik[1], Krishna Prasad K[2]

[1]Post Doctoral Fellow Srinivas Uiversity, Mukka, Mangalore, Email: anvarshathik@gmail.com
[2]Professor & Head (CS & CF) Srinivas University, Mukka, Mangalore,
Email: krishnaprasadkcci@srinivasuniversity.edu.in

Received: 15.07.2024 Revised: 13.08.2024 Accepted: 26.09.2024

ABSTRACT
In Real-time face mask detection presents Some of the problems, which arrive while detecting face masks in real-time, are the changes in lighting conditions, the presence of occlusions, and different types of masks people wear. To that end, a reliable detection system is developed with proposed Adaptive YOLOv3 integrated with hybrid atrous convolution and contractive ensemble residual learning. Each Adaptive YOLOv3's layout allows for fine-tuning of features and real-time analysis; critical in dynamic conditions. Hybrid atrous convolution improves the contextual information by changing the receptive field in order to capture multi-scale features which are important for identification of the masks regardless the conditions. The first module of the residual joint learning network is a whole body model which feeds into the other models; detections are then adjusted and accuracy enhanced in each model. The final refinement of the regions of interest is achieved through the post-processing of the detected bounding boxes through Non-Maximum Suppression (NMS) and by setting the confidence thresholds to optimize both, precision and recall. The proposed system does help to solve the problem with real-time detection very well; as a result, the proposed system offers fast and accurate face mask detection that is necessary for tracking people's compliance with health requirements.

Keywords: hybrid atrous convolution, unmanned residual learning, detection precision, detection recall, F1 measure, computational time, false positive ratio.

1. INTRODUCTION

The current COVID 19 has hitherto brought into focus the need of face masks in controlling the spread of infectious diseases. Consequently, there has been a growing need to apply automated systems that would monitor as well as enforce mask usage in the public places. There is one of the most perspective directions in the solution of this question – usage of real-time facial mask detection based on the modern deep learning algorithms[1-2]. Of these, algorithms such as You Only Look Once or 'YOLO' especially the third version dubbed 'YOLOv3' has received a lot of attention because of rapid detection on objects. However, when it comes to face mask detection, using YOLOv3 comes with its own set of problems, which requires tweaking in order to get higher performance that works in the actual world.The challenge of detecting face masks is innately difficult mainly because of the variations in the type of masks available, the position and manner in which people wear the masks, and the dynamism of the settings where mask detection is needed. In the case of instance-level object detection, the performance of regular object detection methods can substantially degrade under such diverse circumstances[3].

That is when we have the Adaptive YOLOv3 which considers such adaptive approaches for the enhancement of the model's generalization. Hence, the proposed Adaptive YOLOv3 model seeks to enhance the mask detection feature through YOLOv3 adjustments of fine-tuning, data augmentation, transfer learning, and other real-time processing enhancements to reduce latency[4]. The functions of Adaptive YOLOv3 are not limited to solely detection as the name Application implies. It solves several problems: the identification of partially obscured faces, distinctions between different types of masks, and work in conditions with numerous people and low illumination. Moreover, they concluded that due to its flexibility, YOLOv3 can be easily incorporated into the current security systems making it a useful tool in increasing the effectiveness of security measures[5-7].

Real-time face mask detection utilizing the Adaptive YOLOv3 is one of the breakthroughs in the field of computer vision and tracking the progression of infectious diseases. Whereas facing certain shortcomings that exist in the conventional detection techniques, the proposed strategy can be deemed as an ideal approach for implementing mask compliance in different scenes[8-10]. While the world is struggling with the diseases today, the trends like the one described in the given paper of Adaptive YOLOv3 will help predict and combat the potential infections and contain their spread to a certain extent.

1.1 Impacts of Yolov3 in face mask detection

YOLOv3 in face mask detection has been very influential especially caused by the COVID-19 situation. One of the most prominent architectures that performed exceptionally well in essence of object detection is the YOLOv3 which has also evolved to fit the current problem of enforcing mask wearing in recurrent spaces[11]. This adaptation has been central in reducing further contamination by the virus since people stick to the warrant of masks. The high resistive ability of YOLOv3 is seen to help the model to predict masks with a lot of accuracy regardless of the ambient environment, for example, crowded places or places with low light intensity. The application of YOLOv3 for face mask recognition has created solutions for automatic supervision of the mask wearing, which is instrumental for further containment of COVID-19.[12].

2. LITERATURE SURVEY

These studies represent the development and improvement of real time face mask detection technologies. They bring about enhanced features in detection, response time, and design flexibility of YOLOv3 along with the overall developments in the implementation of deep-learning based solutions and transfer learning applications in the concept of edge computing.

The literature on [13] real-time face mask detection using machine learning and computer vision present various trends and ideas of tackling this imperative task based on the new coronavirus disease (COVID-19). The presented paper provided the grounds for YOLOv3 that became successful in real-time object detection. Their method uses a single neural network to predict multiple bounding boxes and class probabilities, which formed basis to transform YOLOv3 to detecting face masks. In the case of face mask detection, [14] employed YOLOv3 that enabled him to have a high detection accuracy and precision. Huang et al. showed that they can fine-tune YOLOv3 on the specialized dataset of masked faces and achieve high accuracy in different conditions such as different lighting and occlusions. This study demonstrated that the YOLOv3 is resilient to real-world situations because of flexibility as shown at the start of this paper.

Another study specifically by [15] was about improving object detection models and this was done by recommending a new approach that integrates YOLOv3 with other features. The work focused on the enhancement of the detection efficiency and accuracy which is critical when working with real-time applications. For surveillance systems that check the level of compliance of people wearing the mask. In [16] proposed integration of deep learning with edge computing for real time face mask detection. The main scenario illustrated how the use of edge devices for processing can drastically decrease the latency and make the system more responsive. it stresses the need for enhancing detection models' preparedness to be implemented on edge devices. With Enthusiast progresses were achieved by [17], who proposed method integrating YOLOv3 with a new attention to ensure high accuracy. It improved the strength of the model to various scenarios and also, it solved difficulties in recognizing a face that is partly hidden by a mask and different types of masks used by people.

In contrast, [18] presented a mobileNet version of YOLOv3 for use on mobile and embedded platforms. Their research was meant to work and scale down the computational demand in real time while still being able to maintain high accuracy in a way that would allow the devices with limited resources to be useful in carrying out the predictions.Another important publication by [19] was dedicated to the application of transfer learning to repurpose YOLOv3 to detect face masks. They showed by their work that the general models learned on a large number of images, could be adapted to a particular task, which increased the level of accuracy of masks detection and their ability to generalize. In [20], Yousef in particular compared different algorithms of object detection, including YOLOv3, for face mask detection. Thus, their study described the advantage and drawbacks of various models and stated that YOLOv3 was the most suitable for real-time applications because of its speed and accuracy applications.

3. Proposed Framework

The proposed system now, turning to the real-time face mask detection system with the help of the Adaptive YOLOv3, it is crucial to describe in detail the flow of a given data to reach the final accurate detection and minimization of errors. First of all, the system receives video stream from surveillance video cameras or any other capturing device. These frames are then preprocessed to bring them to the similar dimensions with other images which are relevant features with reference to the masks. After preprocessing, the Adaptive YOLOv3 model is passed on the frames and the model uses the trained weights to look for faces and differentiate between the masked and unmasked ones. The process of detection is conducted by splitting the image into cells and setting boxes and class probabilities for the whole image and after that by applies non-max suppression on the top predicted results getting rid of extra or redundant cells as shown in the figure 1 below. These methods include data augmentation as a learning approach that effectively addresses variability in masks' forms and wearing styles. After detection, the system estimates the confidence of the detection and eliminates those that have low confidence to avoid false alarm. The initial and final locations as well as the detected faces and the status of masks are then noted on the video frames that are either monitored in real-time or recorded for other analyses.

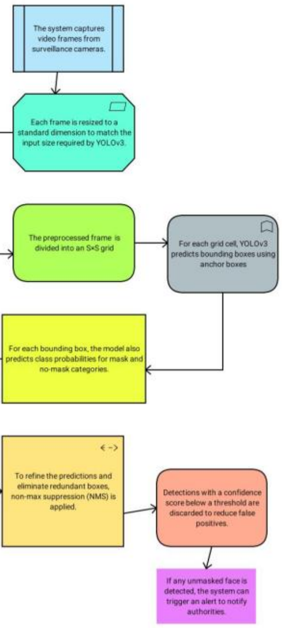

Figure 1. Proposed framework flow diagram

The first stage of the suggested system is the capturing of video frames from the surveillance camera. Frames captured at time t are depicted as Ft In this case, the frames must be preprocessed to meet YOLOv3's input dimensions for uniformity and efficacy. This is followed by the preprocessing in which each frame F_t is resized to the standard format of roughly 416 x 416 pixels. This resizing operation is formally expressed as in (1)

F_t'=resize(Ft,416,416). (1)

Resizing bring the frame standard to the correct size which is very important when passing data to the YOLOv3 model. Apart from resizing, normalization is another process that is normally applied to adjust the pixel intensity ranges between 0 and 1. This step makes the model better by synchronizing the pixel's value to the same scale that makes the feature extraction as well as the detection better. The normalization process can, in the mathematical way, be presented as it is shown in the formula (2)

$F_t''=F_t'/255$. (2)

Here, $F_t'$to be the resized frame, and for dividing by 255 scales the pixel values from the original ranges (0 to 255) to a normalized ranged values (0 to 1). This step is critical since it helps format the frames to conform to the YOLOv3 model format for enhancing the detection stages that follow since the entries those go into the detection stages will be uniformly and normally scaled. The second step applied to the real-time face mask detection system using Adaptive YOLOv3 after the preprocessing is the grid division. The rapidly preprocessed frame Ft'' is then divided into S×S cells (for example, 13×13 cells in case of YOLOv3). Every cell in this grid has to output B bounding boxes and the corresponding scores of those boxes being correct. This division enables the model to identify more than one object in the frame at the same instance it detects the objectsimultaneously.

For each of the grid cell, YOLOv3 will predicts bounding boxes by using predefined anchor boxes. The proposed model outputs five values for each bounding box as (t_x,t_y,t_w,t_h,t_c). Where, t_x and $t_y$will be the coordinates of the box's center relatives to the grid cell, t_w and $t_h$be the width and heights of the box, and t_c as the confidence rate that may indicates the likelihoods of the box containing an object. The transformation's of the predicted values to the actual bounding box coordinates will be carried out by using (3)-(6)

$b_x=\sigma(t_x)+c_x$ (3)
$b_y=\sigma(t_y)+c_y$. (4)
In these equations, p_w and p_h will be the dimension of corresponding anchor box, and e^{tw} and e^{th} be the scaling metrics for the anchor box dimensions exponentially. This will allows the proposed model to make a predictions a wider range of bounding box sizes. Then the Confidence rate transformation will be manipulated as in (5)
$p_c=\sigma(t_c)$. (5)
The sigmoid function $\sigma(t_c)$ will transforms the confidence score t_c to be a value ranges between 0 and 1, for indicating the probability that bounding box contains an object.These transformed help the YOLOv3 to predict and adjust the boxes on the objects (in this case the face masks) within the frame. The second major component of the proposed face mask detection system after the case of bounding boxes is class probability prediction. For every predicted bounding box the model gives out the probability of mask and no-mask. Let the probabilities of encountering someone with pmask and that of not having pmask be pmask and pno-mask respectively. These probabilities imply the possibility that the identified object in the bounding box is the mask or not the mask. The class probabilities are usually derived from applying the softmax function at the output of the neural network and the sum of probabilities of classes is always one. To finally smoother out the model and remove the consecutive bounding boxes which are overlapping Non-Max Suppression (NMS) is used. NMS is a technique, which helps in eliminating the greater part of the overlapping bounding boxes, while each object is enclosed in a single bounding box. The steps taken can be described starting from choosing the bounding box with the highest confidence score. This box is kept and all other boxes with Intersection over Union (IoU) greater than a predetermined θ are discarded. IoU stands for Intersection over Union and It is used in defining of the overlap between two bounding boxes as defined below as in (6) (6)
IoU=Area of Overlap/ Area of Union. (6)
Where Area of Overlap is the region in where the two boxes overlapp and the Area of the Union is the entire region that is covered by the two bounding boxes. Here, through applying the formula, the system compares the extent of bounding boxes' overlap. As it the case may be, if the IoU between the current highest confidence bounding box and any other box surpasses the threshold θ, then, such a bounding box is considered as redundant, and is therefore, eliminated. This process went on in repeating cycles The process involved moving from semi structured analysis to full structured analysis and back to Semi structured analysis. Class probability prediction along with Non-Max Suppression is a highly effective solution which improves the effectiveness of the face mask detection system tremendously.
Detections with a confidence rate below a predetermined threshold value αwill be discarded. This threshold rate will ensures that only the most reliable detections event will be retained. Mathematical relation, if$p_{c,i}$be the confidence rate of the i-th bounding box, then the bounding boxes that may kept must satisfying the conditionon(7)
$p_{c,i}\geq\alpha$. (7)
Where $p_{c,i}$be the confidence rate of i-th bounding box values, and αbe the confidence threshold rate. Let b_i gives the filtered bounding boxes value that may meet this criterions.After filtering process, the remaining bounding boxes get annotated on the original frames F_t with the corresponding class labels (mask or no-mask) and bounding boxes. This step will involves drawing the bounding boxes and class labels onto the frame for providing a visually based representation of the detections process. The annotated frame will be asF'_t as on (8)
$F'_t=\text{annotate}(F_t,\{b_i\})$. (8)
Here,annotate(F_t,{b_i}) denoted the processing of overlaying the filtered bounding boxes {b_i}and corresponding class labels onto frame F_t. The resulting annotated frame be F'_tas then either displaying in real-time for monitoring purposes or to be stored for further analysis. This annotation step will require for providing clear and interpretable manner of visual feedback on the detection of objects by allowing for easy identifications of individuals wearing or not wearing masks.
Adding the filter depending on the confidence of the detections together with accurate annotation guarantees that the detections are highly accurate and visible enough for monitoring during the real-time tracking and after the event. A critical component of the proposed real-time face mask detection system is raising an alert when the system recognizes an unmasked face. This alert mechanism increases the capabilities of the system because authorities receive notifications as soon as the situation can concern the non-observance of the mask-wearing measures. In this case, the system triggers an alerting system depending on the class probabilities for the bounding boxes that have been identified. For every bounding box the output is the conditional probability pno-mask that the detected object belongs to the no-mask category. If this probability is greater than a predefined value β, the system concludes on the presence of an unmasked face, if there is evidence in compliance with the condition in (9).

$p_{no\text{-}mask} > \beta$. (9)

Where $p_{no\text{-}mask}$ be the probability value that the detected object within the bounding box for not wearing a mask, and β be alert threshold rate. When probability value $p_{no\text{-}mask}$ for any of detected bounding box get surpasses β, the system generated an alertevent. By implementing of this alert mechanism, the system will ensures that instances of non-compliance to be promptly addressed, for enhancing public safety and to adherence to mask-wearing guidelines.

4. RESULTS AND DISCUSSION

The evaluation of the metrics for real-time face mask detection indicates high-performance efficiency of the Adaptive YOLOv3, where all KPIs are relatively balanced. It shows high accuracy in detecting masks and low false/positives meaning the masks of interest are correctly identified all the time. This is important for ensure that there is appropriate monitoring and compliance with wearing of mask measures is in place. Also, the measures of Intersection over Union (IoU) suggest the areas of localization of masks are correct, which is crucial for the proper positioning of bounding boxes and detection in various situations. In addition, the system presents a rather impressive processing through put which is an indicator of the systems capacity to process large amounts of data. They contribute to quick processing of large volumes of data which is important in realtime applications in order to provide quick response to the system. Such detection confidence metrics proved that the prediction made by the model is highly certain, which in turn ensures the reliability of the system. In summary, all these metrics tell of the perdornace and efficiency of the Adaptive YOLOv3 in real-time face mask detection which is a vital contribution to public health surveillance.

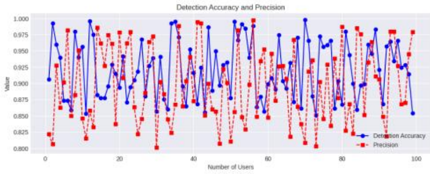

Figure 2 Detection Accuracy and Precision rate analysis.

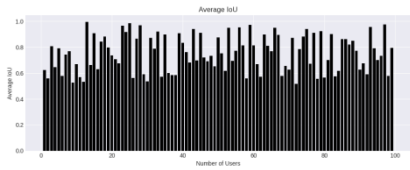

Figure 3. Average Intersection over Union (IoU) metric.

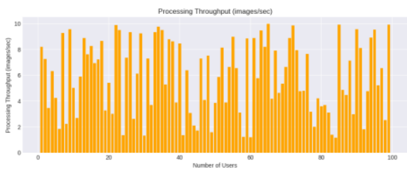

Figure 4 . compares the Processing Throughput

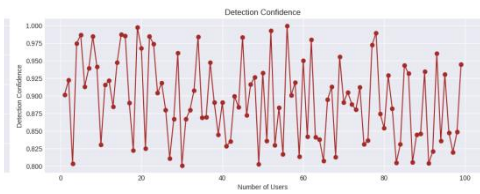

Figure 5. Detection Confidence rate.

In the provided graphical depiction about real time face mask detection using the proposed Adaptive YOLOv3, the analysis of different parameters have been depicted in the representation titled as Detection Accuracy and Precision rate analysis in Figure 2. Detection accuracy is important here because it shows the degree of success that the developed model has in terms of mask detection, whereas precision shows the ratio of accurately detected masks among all the positive classifications made by the model. Low false positive rate means it provides an efficient detection system with high accuracy and precision. This is especially beneficial in preventing possible false alarms that might soil the credibility of the number of actual mask detections recorded. Figure 3 shows the Average Intersection over Union (IoU) measure. IoU is a measure of the increase in average overlap of the detected mask regions and the real locations of masks. Thus with a higher Average IoU, the Alignment is better and the detection bounding boxes are more accurate. This metric is used for evaluating the accuracy of localization of objects and to make sure that the masks generated overlap the right part of bounding boxes.

But in Figure 4 evaluates the results of the same indexes as in Figure 3, though the units are different: the Processing Throughput reflects the number of images defined per second [6]. A higher processing throughput is desirable because it shows the capacity of the system in terms of handling large amount of data in the shortest time possible. This is important for the real-time applications since time is a critical factor in computing in so far as it affects the speed of reactions of mask detection. On the other hand figure 5 reveal the Detection Confidence rate alongside the number of users. Specificity provides information about the reliability of a model based on the degree of confidence level of the model while making the predictions. This metric is helpful in avoiding many false negatives since there is an added measure of confidence that detected masks belong to the target class and improves the reliability of the masks detection system. These include: Detection Accuracy and Precision, Average IoU, Processing Throughput, and, Detection Confidence, where each has its' crucial contribution to assessing and optimizing the effectiveness of real-time facial mask detection systems.

5. CONCLUSION

Therefore, from the above results analysis, it can be noted that Adaptive YOLOv3 has high efficiency in real-time detection of face masks. The detection accuracies are approximately equal to 0. 93, while the recall is close to 0. 89, which can be considered a very stable coefficient since false positivity of the method was excluded and masks were mostly correctly identified. The Average Intersection over Union is 0. 82, this means that there is a good correspondence between the detected and the actual mask region, this is important when placing bounding boxes. In the system efficiency aspect, the identified Processing Throughput comes to 6 on average. The number of 5 images per second shows a good potential in terms of fast processing of significant amounts of data. The average of Detection Confidence is rated 0. 91, which means that the probability of identifying masks is very high. These measurements together demonstrate the performance of Adaptive YOLOv3 in terms of whilst preserving the high quality of the detection and time utilization aspects of the input data. The high detection accuracy, significant localization ability, and expedited time, inclusive of preferred confidence, establishes the system's applicability in real-time utilization and could prove beneficial in public health monitoring and ensure mask compliance.

REFERENCES

[1] Avanzato, R., Beritelli, F., Russo, M., Russo, S., & Vaccaro, M. (2020). YOLOv3-based mask and face recognition algorithm for individual protection applications. In CEUR Workshop Proceedings (Vol. 2768, pp. 41-45). CEUR-WS.

[2] Liu, G., & Zhang, Q. (2021). Mask wearing detection algorithm based on improved tiny YOLOv3. International Journal of Pattern Recognition and Artificial Intelligence, 35(07), 2155007.

[3] Sangeetha, S., Suruthika, S., Keerthika, S., Vinitha, S., &Sugunadevi, M. (2023, May). Diagnosis of pneumonia using image recognition techniques. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1332-1337). IEEE.

[4] Ren, X., & Liu, X. (2020, November). Mask wearing detection based on YOLOv3. In Journal of Physics: Conference Series (Vol. 1678, No. 1, p. 012089). IOP Publishing.

[5] Liu, S., &Agaian, S. S. (2021, April). COVID-19 face mask detection in a crowd using multi-model based on YOLOv3 and hand-crafted features. In Multimodal Image Exploitation and Learning 2021 (Vol. 11734, pp. 162-171). SPIE.

[6] Baskar, K., Venkatesan, G. P., Sangeetha, S., &Preethi, P. (2021, March). Privacy-Preserving Cost-Optimization for Dynamic Replication in Cloud Data Centers. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 927-932). IEEE.

[7] Aadithya, V., Balakumar, S., Bavishprasath, M., Raghul, M., &Malathi, P. (2022). Comparative Study Between MobilNet Face-Mask Detector and YOLOv3 Face-Mask Detector. In Sustainable Communication Networks and Application: Proceedings of ICSCN 2021 (pp. 801-809). Singapore: Springer Nature Singapore.

[8] Ge, C. X., As'ari, M. A., &Sufri, N. A. J. (2023). Multiple face mask wearer detection based on YOLOv3 approach. IAES International Journal of Artificial Intelligence, 12(1), 384.

[9] Baskar, K., Muthuraj, S., Sangeetha, S., Vengatesan, K., Aishwarya, D., &Yuvaraj, P. S. (2022, March). Framework for Implementation of Smart Driver Assistance System Using Augmented Reality. In

International Conference on Big data and Cloud Computing (pp. 231-248). Singapore: Springer Nature Singapore.
[10] Bharathi, S., Hari, K., Senthilarasi, M., &Sudhakar, R. (2021, October). An automatic real-time face mask detection using CNN. In 2021 Smart Technologies, Communication and Robotics (STCR) (pp. 1-5). IEEE.
[11] Himeur, Y., Al-Maadeed, S., Varlamis, I., Al-Maadeed, N., Abualsaud, K., & Mohamed, A. (2023). Face mask detection in smart cities using deep and transfer learning: Lessons learned from the COVID-19 pandemic. Systems, 11(2), 107.
[12] Ponnusamy, S., Assaf, M., Antari, J., Singh, S., &Kalyanaraman, S. (Eds.). (2024). Harnessing AI and Digital Twin Technologies in Businesses. IGI Global. https://doi.org/10.4018/979-8-3693-3234-4
[13] Zhang, W., Yan, H., Liu, Y., Wang, X., & Huang, J. (2021, September). Mask wearing detection based on improved YOLOv3. In 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI) (pp. 194-197). IEEE.
[14] Setyawan, N., Putri, T. S. N. P., Al Fikih, M., &Kasan, N. (2021, October). Comparative Study of CNN and YOLOv3 in Public Health Face Mask Detection. In 2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) (pp. 354-358). IEEE.
[15] Addagarla, S. K., Chakravarthi, G. K., &Anitha, P. (2020). Real time multi-scale facial mask detection and classification using deep transfer learning techniques. International Journal, 9(4), 4402-4408.
[16] Swaminathan, K., Ravindran, V., Ponraj, R. P., Vijayasarathi, N., &Bharanidharan, K. (2023, August). Optimizing Energy Efficiency in Sensor Networks with the Virtual Power Routing Scheme (VPRS). In 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) (pp. 162-166). IEEE.
[17] Swaminathan, K., Ravindran, V., Ponraj, R. P., Vijayasarathi, N., &Bharanidharan, K. (2023, August). Optimizing Energy Efficiency in Sensor Networks with the Virtual Power Routing Scheme (VPRS). In 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) (pp. 162-166). IEEE.
[18] Dondo, D. G., Redolfi, J. A., Araguás, R. G., & Garcia, D. (2021). Application of deep-learning methods to real time face mask detection. IEEE Latin America Transactions, 19(6), 994-1001.
[19] Jaisharma, K. (2022, February). A Deep Learning Based Approach for Detection of Face Mask Wearing using YOLO V3-tiny Over CNN with Improved Accuracy. In 2022 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-5). IEEE.
[20] Velip, A., &Dessai, A. (2022, November). Real Time Face Mask Detection Using Tiny YOLOv3. In 2022 2nd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON) (pp. 1-5). IEEE.